# The use of restricted cubic splines to evaluate nonproportional hazards in Cox regression

Viktor Oskarsson  Andrea Discacciati  Nicola Orsini

Unit of Nutritional Epidemiology and Unit of Biostatistics
Institute of Environmental Medicine
Karolinska Institutet

2015 Nordic and Baltic Stata Users Group meeting

# Introduction I

## Cox regression

- Cox regression is the most common regression method for time-to-event data
- An exposure-outcome association is normally quantified into a single hazard ratio (HR)
- A single HR is only meaningful under proportional hazards (PHs)

# Introduction II

## Common approaches to evaluate non-PHs (in Stata)

- Graphically
  - `sts graph`
  - `stphplot`
  - `stcoxkm`

- Numerically
  - `estat phtest`
  - `, tvc(varname)`

# Introduction III

## Common approaches to handle non-PHs (in Stata)

- Piecewise models
  - `stsplit`
  - `, tvc(varname)`
- Stratified models
  - `, strata(varname)`

# Introduction III

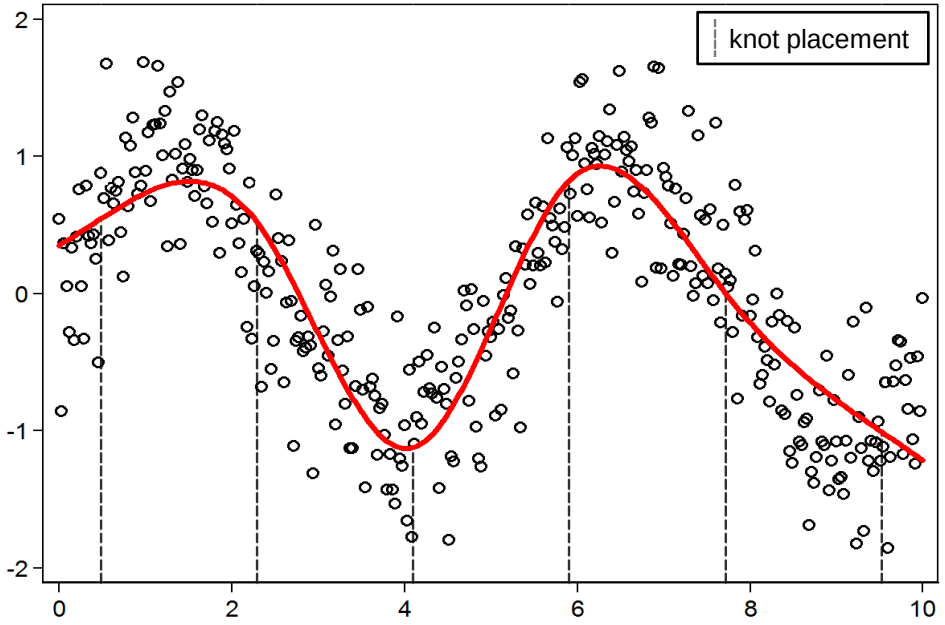
## Common approaches to handle non-PHs (in Stata)

- Piecewise models
  - `stsplit`
  - `, tvc(varname)`
- Stratified models
  - `, strata(varname)`
- Flexible functions
  - Restricted cubic splines (RCS)
  - Fractional polynomials

# Introduction IV

## Cubic splines

- Cubic splines are piecewise-polynomial line segments joined together at knots
- RCS are constrained to be linear before the first knot and/or after the final knot



# Aim of the presentation

- Introduce how RCS can be used in the evaluation of non-PHs in Cox regression
- Present a new Stata postestimation command (`stphcoxrcs`) that greatly facilitates such an evaluation (both numerically and graphically)
- Show a practical example of how to evaluate non-PHs using `stphcoxrcs`

# How to use RCS in the evaluation of non-PHs I

## Preestimation

- First, the analysis time (_*t*) is split into narrow intervals (using `stsplit`)
- Second, a flexible function of _*t* is modeled with RCS (using `mkspline`)
- Third, an interaction is formed between the flexible function of _*t* and the covariate of interest

# Splitting of _t

- Before splitting

| idkod | _d | _t0 | _t |
|---|---|---|---|
| 23011236 | 1 | 74.318256 | 76.538714 |

- After splitting

| idkod | _d | _t0 | _t |
|---|---|---|---|
| 23011236 | 0 | 74.318256 | 74.5 |
| 23011236 | 0 | 74.5 | 75 |
| 23011236 | 0 | 75 | 75.5 |
| 23011236 | 0 | 75.5 | 76 |
| 23011236 | 0 | 76 | 76.5 |
| 23011236 | 1 | 76.5 | 76.538714 |

# How to use RCS in the evaluation of non-PHs I

## Preestimation (cont.)

- First, the analysis time (_*t*) is split into narrow intervals (using `stsplit`)
- Second, a flexible function of _*t* is modeled with RCS (using `mkspline`)
- Third, an interaction is formed between the flexible function of _*t* and the covariate of interest

# How to use RCS in the evaluation of non-PHs II

## Estimation

- A full Cox model is fitted (including the aforementioned interaction terms) (using `stcox`)

$$\lambda(t \mid \mathrm{X}) = \lambda_0(t)\exp(\beta_1 \mathrm{X} + \beta_2 \mathrm{X} \times \mathrm{S}_1(t) + \beta_3 \mathrm{X} \times \mathrm{S}_2(t))$$

# How to use RCS in the evaluation of non-PHs III

## Postestimation

- A joint test of the interaction terms is performed to numerically evaluate non-PHs

$$H_0: \beta_2 = \beta_3 = 0$$

- A graph of the time-varying HRs is created to graphically evaluate non-PHs

$$HR(t) = \exp(\beta_1 + \beta_2 \times S_1(t) + \beta_3 \times S_2(t))$$

# stphcoxrcs

`stphcoxrcs` tests the PH-assumption for a covariate (binary or continuous) using RCS after fitting a model with `stcox`

- *_t* is split using `stsplit` with the options `splitevery(#) | splitat(numlist)`. Default is `splitat(failures)`
- The natural logarithm of *_t* is modeled with RCS using `mkspline` with the option `nknots(#)`. Default is `nknots(3)`

# Practical example I

## Setting and research question

- Data from a prospective cohort of Swedish women (*n* = 32,038; followed-up between 1998-2011)

- Cox regression on the association of fruit and vegetable consumption with risk of symptomatic gallstone disease (*n* = 1120)

- Fruit and vegetable consumption categorized into quartiles of consumption (abbreviated as “fv_cat” in subsequent slides)

# Practical example II

## Cox regression

- `stset`
  - Attained age as time scale

- `stcox`
  - Multivariable-adjusted for potential confounders (omitted from output for clarity)

| _t | Haz. Ratio | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| fv_cat | | | | | | |
| 2 | 0.96 | 0.08 | -0.47 | 0.6388 | 0.81 | 1.14 |
| 3 | 0.93 | 0.08 | -0.81 | 0.4194 | 0.78 | 1.11 |
| 4 | 0.97 | 0.09 | -0.33 | 0.7388 | 0.81 | 1.16 |

# Practical example II

## Cox regression (cont.)

- `stset`
  - Attained age as time scale

- `stcox`
  - Multivariable-adjusted for potential confounders (omitted from output for clarity)
  - Postestimation: `estat phtest, detail`

| | rho | chi2 | df | Prob>chi2 |
|---|---|---|---|---|
| 1b.fv_cat | . | . | 1 | . |
| 2.fv_cat | 0.01749 | 0.34 | 1 | 0.5603 |
| 3.fv_cat | 0.01998 | 0.44 | 1 | 0.5074 |
| 4.fv_cat | 0.07726 | 6.65 | 1 | 0.0099 |

# Practical example III

## Evaluation of non-PHs using `stphcoxrcs`—test statistics

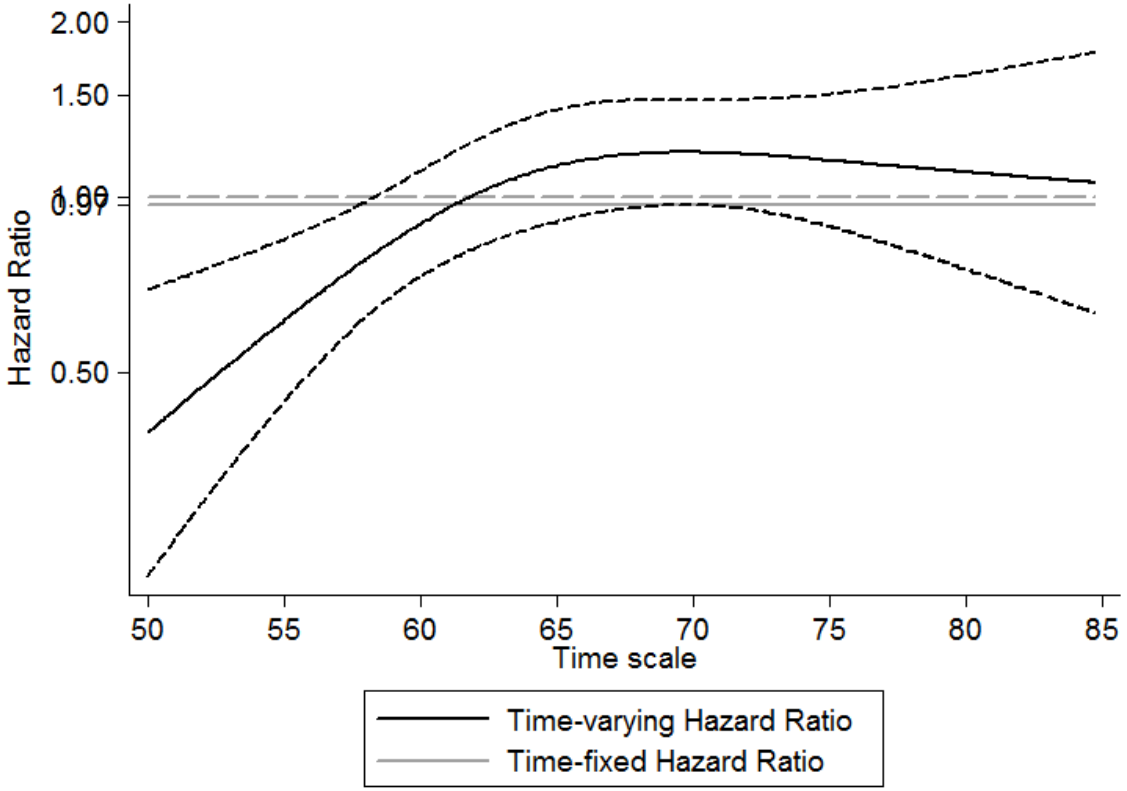
- `stphcoxrcs 4.fv_cat`

```
Wald test of proportional-hazards assumption for i4.fv_cat

    chi2(2) = 12.29
    Prob > chi2 = 0.0021

    Note: time scale modeled using Restricted Cubic Splines with 3 knots
```

# Practical example IV

## Evaluation of non-PHs using `stphcoxrcs`—graphics

- `stphcoxrcs 4.fv_cat`



# Practical example V

## Evaluation of non-PHs using `stphcoxrcs`—options

- Main
  - `nknots(#)`
  - `splitevery(#)`
  - `splitat(numlist)`

- Test statistics
  - `ic`
  - `lrtest`

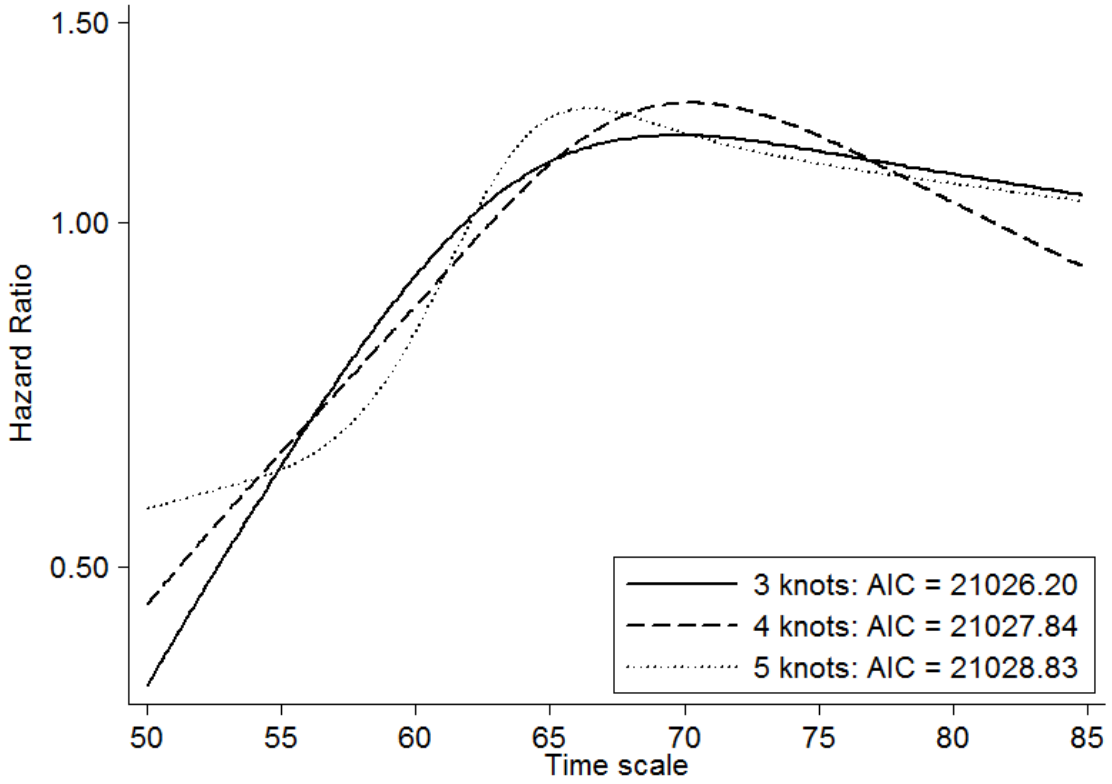
# Practical example V

## Evaluation of non-PHs using `stphcoxrcs`—options (cont.)

- Graphics
  - `level(#)`
  - `range(# #)`
  - `noci`
  - `noyref`
  - `nograph`
  - `gopts(twoway_options)`
  - `saving(filename [, replace])`

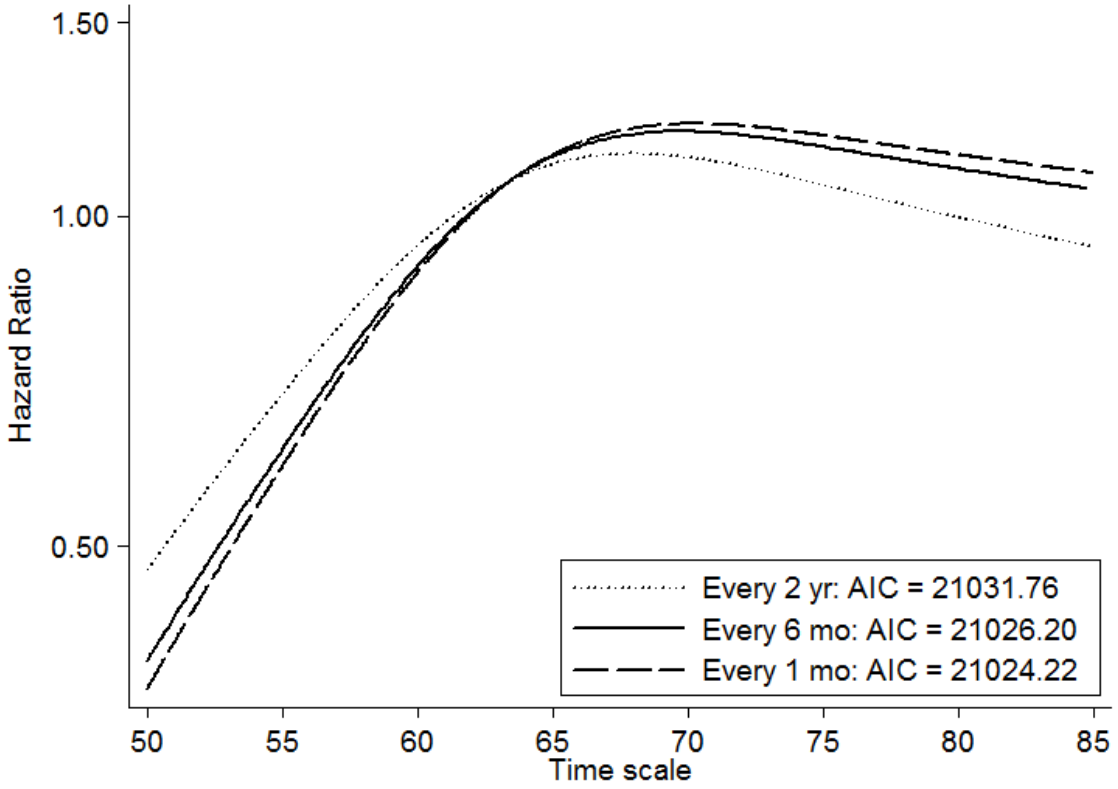
# Practical example VI

Comparison of Cox models with different number of RCS-knots using `stphcoxrcs`



# Practical example VII

Comparison of Cox models with different intervals of *_t* using `stphcoxrcs`



# Summary

- In Cox regression, the assessment of PHs is important for the interpretation of HRs and exposure-outcome associations
- `stphcoxrcs` facilitates the use of RCS in the evaluation of non-PHs (both numerically and graphically)
- It is a useful complement to other approaches that are used to evaluate and/or handle non-PHs

# References


- Heinzl, H., & Kaider, A. (1997). Gaining more flexibility in Cox proportional hazards regression models with cubic spline functions. Computer methods and programs in biomedicine, 54(3), 201-208.

- Therneau, T. M., & Grambsch, P. M. (2000). Modeling survival data: extending the Cox model. Springer-Verlag. New York.

- Royston, P., & Lambert, P. C. (2011). Flexible parametric survival analysis using Stata: beyond the Cox model. Stata Press books.

# Practical example II

## Cox regression (cont.)